

Nedílnou součástí vývoje léčiv je tzv. virtuální screening, jehož cílem je počítačová identifikace biologicky aktivních molekul. Jednou z variant virtuálního screeningu je ligandový virtuální screening, jenž je založen na využití známých biologicky aktivních molekul a podobnostního vyhledávání. Molekulu lze reprezentovat jako graf, molekulární podobnost lze pak modelovat na základě stejných fragmentů (podgrafů) mezi dvěma molekulami. Běžnou praxí je fragmenty hashovat do omezeného číselného intervalu a používat tato hashovaná čísla pro výpočet molekulární podobnosti. Při tomto hashování ovšem může dojít ke kolizím. Obecně jsou kolize považovány za nežádoucí, neb dochází ke ztrátě informace o molekule. Naším cílem bylo vyzkoušet, zda-li mohou kolize fragmentů vést k lepším výsledkům. Za tímto účelem jsme navrhli několik podobnostních modelů postavených na fragmentech. Pro účely vyhodnocení jsme implementovali testovací prostředí, jenž umožňuje snadné testování a vyhodnocení různých modelů. Z provedených experimentů plyne, že vybrané kolize vedou k lepším výsledkům, než jsou výsledky běžně používaných metod. Dokonce existují kolize, které v určitém modelu dosahují AUC přesahující 0.99.